

Inferência Estatística III

Análises de Variâncias

Prof. Dr. Juliano van Melis

Parte I

Parte 1

Conteúdo

• Análise de Variância – ANOVA

- Introdução
- Pressupostos da ANOVA
- Teste de Levine para homogeneidade da variância
- Estatística F para testar igualdade de várias médias
- Interpretação do Quadro ANOVA
 - Outputs R, SPSS e SAS
- ANOVA de um fator com o MS Excel® e R
- ANOVA com dois fatores com MS Excel® e R
- ANOVA com medidas repetidas

• Teste de Kruskal-Wallis

Parte 2

Conteúdo

• Correlação Linear Simples

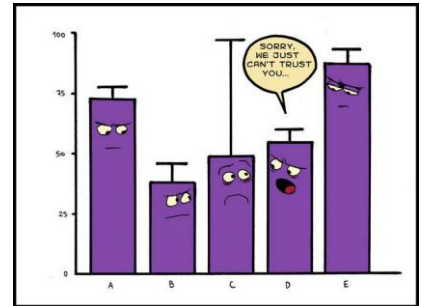
- Coeficiente Correlação Linear de Pearson
- Significância da correlação linear
- Medida de associação paramétrica
- Teste t student para análise da significância CLP
- Aplicações e análises com MS Excel® e R

• Medida de associação não-paramétrica

- Teste de Spearman

• Correlação Bisserial

• Avaliação

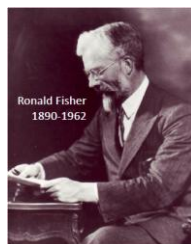


Análises de Variâncias

ANOVA - ANÁLISE DE VARIÂNCIA

Introdução & Pressupostos Homocedasticidade & Médias ANOVA de um fator ANOVA com dois fatores ANOVA medidas repetidas

→ Teste paramétrico (possui estimativas de parâmetros)
 → Utilizado quando o pesquisador deseja verificar se existem diferenças entre as **médias** de uma **determinada variável (variável resposta)** em relação a um **tratamento** com dois ou mais níveis **categóricos (variável preditora)**.
 → O teste t também é utilizado nesse tipo de procedimento (com no máximo de dois níveis), porém a ANOVA é indicada para análises com $n > 30$.

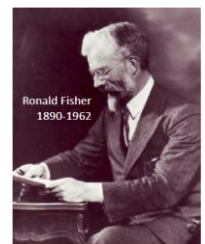


ANOVA
Maximum likelihood estimation

Variável resposta (métrica) ~ Variável preditora (categórica)

Introdução & Pressupostos Homocedasticidade & Médias ANOVA de um fator ANOVA com dois fatores ANOVA medidas repetidas

Definição: "Técnica estatística usada para determinar se as amostras de dois ou mais grupos surgem de populações com médias iguais. A análise de variância é empregada para uma medida dependente"



ANOVA
Maximum likelihood estimation

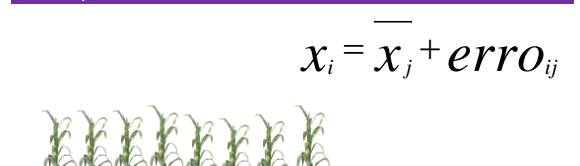
Variável Dependente (métrica) ~ Variável Independente (categórica)

Introdução & Pressupostos Homocedasticidade & Médias ANOVA de um fator ANOVA com dois fatores ANOVA medidas repetidas



$$\bar{x}_j = \frac{\sum_{n=1}^i x_n}{i}$$

Introdução & Pressupostos Homocedasticidade & Médias ANOVA de um fator ANOVA com dois fatores ANOVA medidas repetidas



$$x_i = \bar{x}_j + erro_{ij}$$

Plantação j

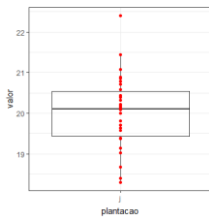
$$erro_{ij} = x_i - \bar{x}_j$$

Introdução & Pressupostos Homocedasticidade & Médias ANOVA de um fator ANOVA com dois fatores ANOVA medidas repetidas

$$x_i = \bar{x}_j + erro_{ij}$$

```
dados <- data.frame(valor = 20 + rnorm(30),
  plantacao = rep("j", 30))
```

```
ggplot(dados,
  aes(x=plantacao,
    y=valor)) +
  geom_boxplot(alpha=0.4) +
  geom_point(color='red') + theme_bw()
```



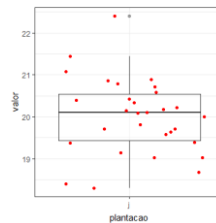
ERRO "dentro" do grupo (within)

Introdução & Pressupostos Homocedasticidade & Médias ANOVA de um fator ANOVA com dois fatores ANOVA medidas repetidas

$$x_i = \bar{x}_j + erro_{ij}$$

```
dados <- data.frame(valor = 20 + rnorm(30),
  plantacao = rep("j", 30))
```

```
ggplot(dados,
  aes(x=plantacao,
    y=valor)) +
  geom_boxplot(alpha=0.4) +
  geom_jitter(color='red') + theme_bw()
```



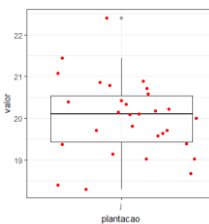
$$SSD_W = \sum_i \sum_j (x_{ij} - \bar{x}_i)^2$$

Introdução & Pressupostos Homocedasticidade & Médias ANOVA de um fator ANOVA com dois fatores ANOVA medidas repetidas

$$x_i = \bar{x}_j + erro_{ij}$$

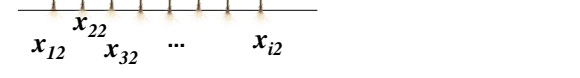
```
dados <- data.frame(valor = 20 + rnorm(30),
  plantacao = rep("j", 30))
```

```
ggplot(dados,
  aes(x=plantacao,
    y=valor)) +
  geom_boxplot(alpha=0.4) +
  geom_jitter(color='red') + theme_bw()
```

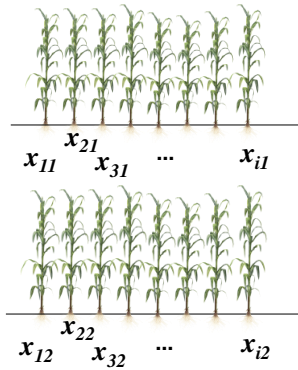


- Média igual a 20
- Erro com média igual a 0 e desvio-padrão igual a 1 (default da função `rnorm()`) para 30 elementos

Introdução & Pressupostos Homocedasticidade & Médias ANOVA de um fator ANOVA com dois fatores ANOVA medidas repetidas



Introdução & Pressupostos Homocedasticidade & Médias ANOVA de um fator ANOVA com dois fatores ANOVA medidas repetidas



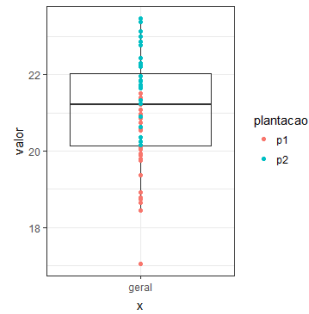
$$x_{i1} = \bar{x}_1 + erro_{i1}$$

$$\bar{x}_1$$

$$\bar{x}_2$$

$$x_{i2} = \bar{x}_2 + erro_{i2}$$

Introdução & Pressupostos Homocedasticidade & Médias ANOVA de um fator ANOVA com dois fatores ANOVA medidas repetidas

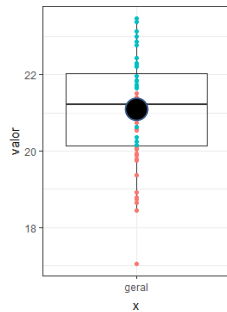


$$x_{i1} = \bar{x}_1 + erro_{i1}$$

$$\bar{x}_* = \frac{\sum x_i}{n}$$

$$x_{i2} = \bar{x}_2 + erro_{i2}$$

Introdução & Pressupostos Homocedasticidade & Médias ANOVA de um fator ANOVA com dois fatores ANOVA medidas repetidas

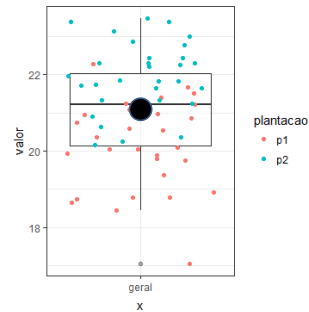


plantacao
p1
p2

$$\bar{x}_* = \frac{\sum x_i}{n}$$

Média geral de todas as observações
(Grande Média)

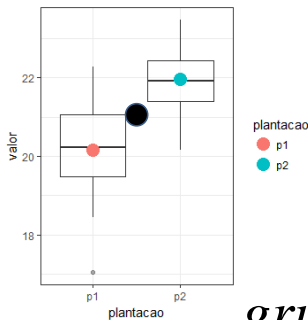
Introdução & Pressupostos Homocedasticidade & Médias ANOVA de um fator ANOVA com dois fatores ANOVA medidas repetidas



plantacao
p1
p2

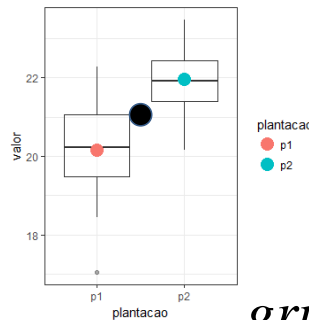
$$\bar{x}_* = \frac{\sum x_i}{n}$$

Introdução & Pressupostos Homocedasticidade & Médias ANOVA de um fator ANOVA com dois fatores ANOVA medidas repetidas



$$grupo_{.j} = \bar{x}_i - \bar{x}_*$$

Introdução & Pressupostos Homocedasticidade & Médias ANOVA de um fator ANOVA com dois fatores ANOVA medidas repetidas



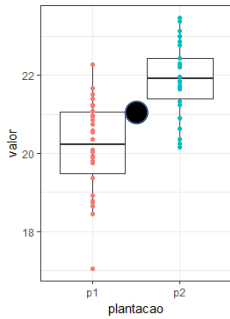
plantacao
p1
p2

$$grupo_{.j} = \bar{x}_i - \bar{x}_*$$

grupo_{ij}:

Erro "entre" os grupos (*between*)

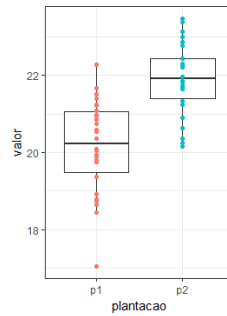
Introdução & Pressupostos Homocedasticidade & Médias ANOVA de um fator ANOVA com dois fatores ANOVA medidas repetidas



$$SSD_B = \sum_i \sum_j (\bar{x}_i - \bar{x}_{..})^2 = \sum_i n_i (\bar{x}_i - \bar{x}_{..})^2$$

$$grupo_{.j} = \bar{x}_i - \bar{x}_{..}$$

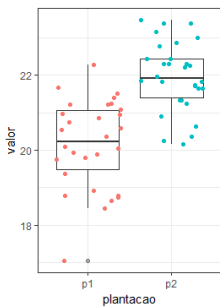
Introdução & Pressupostos Homocedasticidade & Médias ANOVA de um fator ANOVA com dois fatores ANOVA medidas repetidas



$$x_{i1} = \bar{x}_{..} + erro_{i1} + grupo_{.1}$$

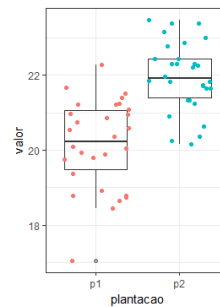
$$x_{i2} = \bar{x}_{..} + erro_{i2} + grupo_{.2}$$

Introdução & Pressupostos Homocedasticidade & Médias ANOVA de um fator ANOVA com dois fatores ANOVA medidas repetidas



$$x_{ij} = \bar{x}_{..} + erro_{ij} + grupo_{.j}$$

Introdução & Pressupostos Homocedasticidade & Médias ANOVA de um fator ANOVA com dois fatores ANOVA medidas repetidas



$$x_{ij} = \bar{x}_{..} + erro_{ij} + grupo_{.j}$$

erro "DENTRO" do grupo j erro "ENTRE" os grupos

Introdução & Pressupostos Homocedasticidade & Médias ANOVA de um fator ANOVA com dois fatores ANOVA medidas repetidas

$$x_{ij} = \bar{x}_{..} + \underbrace{(\bar{x}_i - \bar{x}_{..})}_{\text{erro "ENTRE" os grupos}} + \underbrace{(x_{ij} - \bar{x}_i)}_{\text{erro "DENTRO" do grupo j}}$$

Introdução & Pressupostos Homocedasticidade & Médias ANOVA de um fator ANOVA com dois fatores ANOVA medidas repetidas

Soma dos Desvios Quadrados ENTRE os grupos
"Sum of Squares **BETWEEN**"

$$SSD_B = \sum_i \sum_j (\bar{x}_i - \bar{x}_{..})^2 = \sum_i n_i (\bar{x}_i - \bar{x}_{..})^2$$

Soma dos Desvios Quadrados DENTRO dos grupos
"Sum of Squares **WITHIN**"

$$SSD_W = \sum_i \sum_j (x_{ij} - \bar{x}_i)^2$$

Introdução & Pressupostos Homocedasticidade & Médias ANOVA de um fator ANOVA com dois fatores ANOVA medidas repetidas

É possível normalizar a soma dos quadrados, calculando a **média dos desvios quadrados**

$$MS_W = SSD_W / (N - k)$$

$$MS_B = SSD_B / (k - 1)$$

Mean Squares *Within* & *Between*

N : número total

k : número de grupos

Introdução & Pressupostos Homocedasticidade & Médias ANOVA de um fator ANOVA com dois fatores ANOVA medidas repetidas

$$F = MS_B / MS_W$$

Se o valor de $F = 1$

→ Médias dos Quadrados **ENTRE** os grupos é semelhante às Médias dos Quadrados **DENTRO** dos grupos.

Se o valor de $F < 1$

→ Médias dos Quadrados **ENTRE** os grupos é menor que as Médias dos Quadrados **DENTRO** dos grupos.

→ Nesses dois casos, as variâncias dentro dos grupos são tão grandes que sobressaem a qualquer sinal que os grupos tenham.

Se o valor de $F > 1$

→ Médias dos Quadrados **ENTRE** os grupos é maior que as Médias dos Quadrados **DENTRO** dos grupos.

→ Nesse caso, os grupos parecem ter papel importante para a variação dos valores.

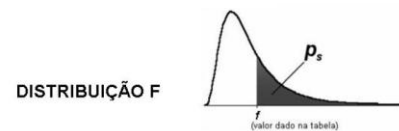
Introdução & Pressupostos Homocedasticidade & Médias ANOVA de um fator ANOVA com dois fatores ANOVA medidas repetidas

Pressupostos

1. **Resíduos (erros)** devem seguir uma distribuição normal: $erro_{ij} \sim N(0, \sigma^2)$
2. **Homogeneidade das variâncias:** As contribuições das variâncias dos grupos devem ser equivalentes para a variância total.
3. **Amostras independentes:** a observação de uma variável não pode influenciar outra observação. Atenção para medidas repetidas!

Introdução & Pressupostos Homocedasticidade & Médias ANOVA de um fator ANOVA com dois fatores ANOVA medidas repetidas

$$F = MS_B / MS_W$$



		Degrees of freedom in numerator (df1)									
	p	1	2	3	4	5	6	7	8	12	24
1	0.100	16.01	10.00	8.01	7.01	6.59	6.35	6.19	6.08	5.95	5.80
	0.050	18.51	12.59	10.13	9.00	8.54	8.29	8.12	8.00	7.87	7.71
	0.025	21.01	14.69	11.81	10.55	10.05	9.78	9.60	9.47	9.33	9.17
	0.010	24.01	17.00	13.78	12.36	11.81	11.50	11.31	11.16	11.00	10.83
	0.001	31.00	22.99	18.51	16.69	16.01	15.67	15.45	15.27	15.09	14.90
2	0.100	19.00	12.00	9.00	8.00	7.59	7.35	7.19	7.08	6.95	6.80
	0.050	21.59	14.18	11.18	10.05	9.59	9.34	9.17	9.05	8.92	8.76
	0.025	24.09	16.59	13.28	12.01	11.54	11.27	11.09	10.96	10.81	10.64
	0.010	27.09	19.00	15.44	14.05	13.56	13.27	13.08	12.94	12.78	12.60
	0.001	34.00	24.99	19.51	17.69	16.99	16.64	16.41	16.22	16.03	15.84

Introdução & Pressupostos Homocedasticidade & Médias ANOVA de um fator ANOVA com dois fatores ANOVA medidas repetidas

→ Teste de Fisher

```
> var.test(valor ~ plantacao, dados)
F test to compare two variances

data: valor by plantacao
F = 0.8676, num df = 29, denom df = 29, p-value = 0.7047
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.412947 1.822824
sample estimates:
ratio of variances
 0.8675999
```

Hipótese nula: Variâncias são iguais

Hipótese alternativa: Variâncias não são iguais

$$W = S_1^2 / S_2^2 \sim F(n - 1, m - 1)$$

Introdução & Pressupostos Homocedasticidade & Médias ANOVA de um fator ANOVA com dois fatores ANOVA medidas repetidas

→ Teste de Fisher

```
> variancias <- tapply(dados$valor, dados$plantacao, FUN=var)
> variancias
      p1      p2 
1.097398 1.264867 
> variancias[1]/variancias[2]
      p1 
0.8675999 
> var.test(valor ~ plantacao, dados)
```

F test to compare two variances

```
data: valor by plantacao
F = 0.8676, num df = 29, denom df = 29, p-value = 0.7047
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.412947 1.822824
sample estimates:
ratio of variances
 0.8675999
```

Introdução & Pressupostos Homocedasticidade & Médias ANOVA de um fator ANOVA com dois fatores ANOVA medidas repetidas

→ Teste de Levene

$W \sim \text{Distribuição F}$
 $gl_1 = k-1$ e $gl_2 = N-k$

Definition [edit]

The test statistic, W , is defined as follows:

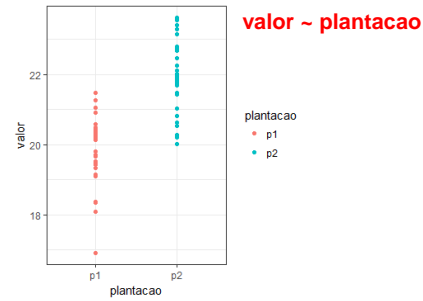
$$W = \frac{(N-k) \sum_{i=1}^k N_i (Z_i - Z_{..})^2}{(k-1) \sum_{i=1}^k \sum_{j=1}^{N_i} (Z_{ij} - Z_i)^2},$$

where

- k is the number of different groups to which the sampled cases belong,
- N_i is the number of cases in the i -th group,
- N is the total number of cases in all groups,
- Y_{ij} is the value of the measured variable for the j -th case from the i -th group,
- $Z_{ij} = \begin{cases} |Y_{ij} - \bar{Y}_i|, & \bar{Y}_i \text{ is a mean of the } i\text{-th group,} \\ |Y_{ij} - \bar{Y}_i|, & \bar{Y}_i \text{ is a median of the } i\text{-th group.} \end{cases}$

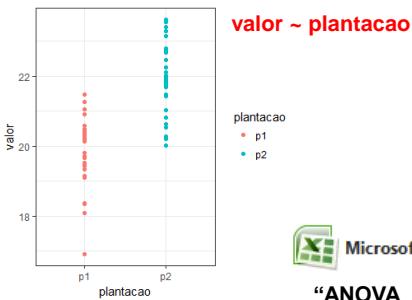
Introdução & Pressupostos Homocedasticidade & Médias ANOVA de um fator ANOVA com dois fatores ANOVA medidas repetidas

EXEMPLO



Introdução & Pressupostos Homocedasticidade & Médias ANOVA de um fator ANOVA com dois fatores ANOVA medidas repetidas

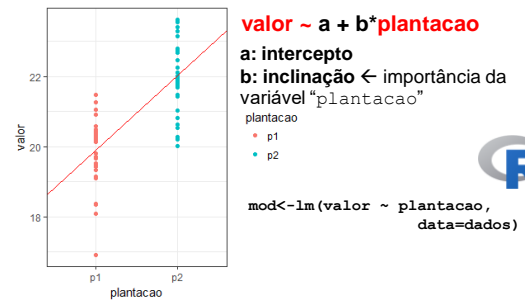
EXEMPLO



Microsoft Office Excel
 “ANOVA_ex01.xls”

Introdução & Pressupostos Homocedasticidade & Médias ANOVA de um fator ANOVA com dois fatores ANOVA medidas repetidas

EXEMPLO



Introdução & Pressupostos Homocedasticidade & Médias ANOVA de um fator ANOVA com dois fatores ANOVA medidas repetidas

EXEMPLO

```
> mod <- lm(valor ~ plantacao, dados)
> mod
```

```
> anova_mod <- aov(mod)
> anova_mod
```

```
> summary(anova_mod)
```



Introdução & Pressupostos Homocedasticidade & Médias ANOVA de um fator ANOVA com dois fatores ANOVA medidas repetidas

EXEMPLO

```
> mod <- lm(valor ~ plantacao, dados)
> mod

Call:
lm(formula = valor ~ plantacao, data = dados)

Coefficients:
(Intercept)  plantacao2
20.183      2.024

> anova_mod <- aov(mod)
> anova_mod
Call:
aov(formula = mod)

Terms:
          plantacao Residuals
Sum of Squares  61.45330  68.50568
Deg. of Freedom      1      58

Residual standard error: 1.086799
Estimated effects may be unbalanced
> summary(anova_mod)

              Df Sum Sq Mean Sq F value    Pr(>F)    
plantacao     1  61.45   61.45    52.03 1.29e-09 ***
Residuals    58  68.51    1.18               

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> |
```



Introdução & Pressupostos Homocedasticidade & Médias ANOVA de um fator ANOVA com dois fatores ANOVA medidas repetidas

EXEMPLO

```
> mod<-lm(valor ~ plantacao, dados)
> mod

Call:
lm(formula = valor ~ plantacao, data = dados)

Coefficients:
(Intercept)  plantacaop2
      20.183      2.024

> anova_mod<-aov(mod)

anova_mod
Call:
aov(formula = mod)

Terms:
      plantacao Residuals
Sum of Squares  61.45330  68.50568
Deg. of Freedom      1         58

Residual standard error: 1.086799
Estimated effects may be unbalanced

> summary(anova_mod)

          Df Sum Sq Mean Sq F value    Pr(>F)    
plantacao   1  61.45   61.45    52.03 1.29e-09 ***
Residuals  58  68.51    1.18               

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> |
```

SSDb = 61.45330

SSDw = 68.50568

Introdução & Pressupostos Homocedasticidade & Médias ANOVA de um fator ANOVA com dois fatores ANOVA medidas repetidas

EXEMPLO

```
> mod<-lm(valor ~ plantacao, dados)
> mod

Call:
lm(formula = valor ~ plantacao, data = dados)

Coefficients:
(Intercept)  plantacaop2
      20.183      2.024

> anova_mod<-aov(mod)

anova_mod
Call:
aov(formula = mod)

Terms:
      plantacao Residuals
Sum of Squares  61.45330  68.50568
Deg. of Freedom      1         58

Residual standard error: 1.086799
Estimated effects may be unbalanced

> summary(anova_mod)

          Df Sum Sq Mean Sq F value    Pr(>F)    
plantacao   1  61.45   61.45    52.03 1.29e-09 ***
Residuals  58  68.51    1.18               

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> |
```

$$F = MS_B / MS_W$$

$$MS_b = 61.45 / 1 = 61.45$$

$$MS_w = 68.51 / 58 = 1.18$$

$$F = 61.45 / 1.18$$

Introdução & Pressupostos Homocedasticidade & Médias ANOVA de um fator ANOVA com dois fatores ANOVA medidas repetidas

EXEMPLO

```
> anova(mod)
Analysis of Variance Table

Response: valor
          Df Sum Sq Mean Sq F value    Pr(>F)    
plantacao   1  61.453   61.453   52.029 1.285e-09 ***
Residuals  58  68.506    1.181               

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> |
```

Introdução & Pressupostos Homocedasticidade & Médias ANOVA de um fator ANOVA com dois fatores ANOVA medidas repetidas

EXEMPLO

```
> mod<-lm(valor ~ plantacao, dados)
> mod

Call:
lm(formula = valor ~ plantacao, data = dados)

Coefficients:
(Intercept)  plantacaop2
      20.183      2.024

> anova_mod<-aov(mod)

anova_mod
Call:
aov(formula = mod)

Terms:
      plantacao Residuals
Sum of Squares  61.45330  68.50568
Deg. of Freedom      1         58

Residual standard error: 1.086799
Estimated effects may be unbalanced

> summary(anova_mod)

          Df Sum Sq Mean Sq F value    Pr(>F)    
plantacao   1  61.45   61.45    52.03 1.29e-09 ***
Residuals  58  68.51    1.18               

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> |
```

$$\text{valor} \sim a + b \cdot \text{plantacao}$$

- O **intercepto** é a Média do 1º Grupo
- A **inclinação** (neste caso) é a diferença do 1º grupo com o grupo seguinte (plantacaop2)

Introdução & Pressupostos Homocedasticidade & Médias ANOVA de um fator ANOVA com dois fatores ANOVA medidas repetidas

TABELA ANOVA 1-FATOR

Fonte da Variação	SQ	gl	Variancia	Razão F
Entre	SQE	k - 1	$S^2_{\text{entre}} = \frac{SQE}{k - 1}$	$F = \frac{S^2_{\text{entre}}}{S^2_{\text{dentro}}}$
Dentro	SQD	n - k	$S^2_{\text{dentro}} = \frac{SQD}{n - k}$	
Total	SQT = SQE+SQD	n - 1		

Introdução & Pressupostos Homocedasticidade & Médias ANOVA de um fator ANOVA com dois fatores ANOVA medidas repetidas



THE POWER TO KNOW.

```
proc anova data = Clover;
  class strain;
  model Nitrogen = Strain;
run;
```

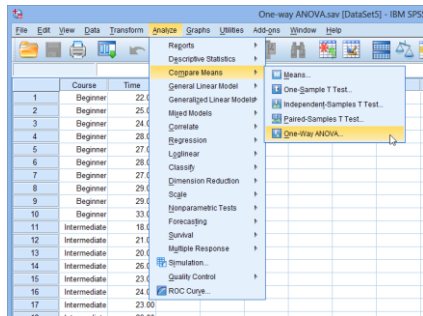
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	847.046667	169.409333	14.37	<.0001
Error	24	282.928000	11.788667		
Corrected Total	29	1129.974667			

R-Square	Coeff Var	Root MSE	Nitrogen Mean
0.749616	17.26515	3.433463	19.88667

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Strain	5	847.0466667	169.4093333	14.37	<.0001

https://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_anova_sec003.htm

Introdução & Pressupostos Homocedasticidade & Médias ANOVA de um fator ANOVA com dois fatores ANOVA medidas repetidas



<https://statistics.laerd.com/spss-tutorials/one-way-anova-using-spss-statistics.php>

Introdução & Pressupostos Homocedasticidade & Médias ANOVA de um fator ANOVA com dois fatores ANOVA medidas repetidas



Time

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Beginner	10	27.2000	3.04777	.96379	25.0168	29.3802	22.00	33.00
Intermediate	10	23.6000	3.30656	1.04563	21.2346	25.9654	18.00	29.00
Advanced	10	23.4000	3.23866	1.02415	21.0832	25.7168	18.00	29.00
Total	30	24.7333	3.58161	.65026	23.4034	26.0633	18.00	33.00

ANOVA

Time	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	91.467	2	45.733	4.467	.021
Within Groups	276.400	27	10.237		
Total	367.867	29			

<https://statistics.laerd.com/spss-tutorials/one-way-anova-using-spss-statistics.php>

Introdução & Pressupostos Homocedasticidade & Médias ANOVA de um fator ANOVA com dois fatores ANOVA medidas repetidas

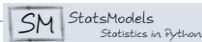


```
from scipy import stats
F, p = stats.f_oneway(d_data['ctrl1'], d_data['trt1'], d_data['trt2'])
```

```
import statsmodels.api as sm
from statsmodels.formula.api import ols

mod = ols('weight ~ group',
           data=data).fit()

aov_table = sm.stats.anova_lm(mod, typ=2)
print aov_table
```



<https://www.marsja.se/four-ways-to-conduct-one-way-anovas-using-python/>

Introdução & Pressupostos Homocedasticidade & Médias ANOVA de um fator ANOVA com dois fatores ANOVA medidas repetidas

EXEMPLO



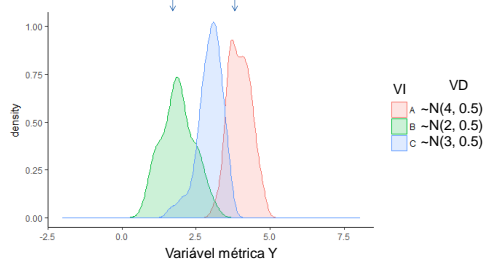
```
> t.test(valor ~ plantacao, dados)
```

Welch Two Sample t-test

```
data: valor by plantacao
t = -7.2131, df = 57.71, p-value = 1.318e-09
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2.585840 -1.462314
sample estimates:
mean in group p1 mean in group p2
 20.18286      22.20693
```

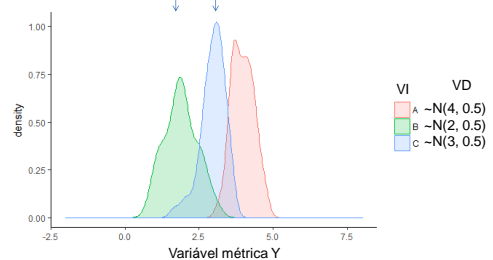
Introdução & Pressupostos Homocedasticidade & Médias ANOVA de um fator ANOVA com dois fatores ANOVA medidas repetidas

Teste t 1

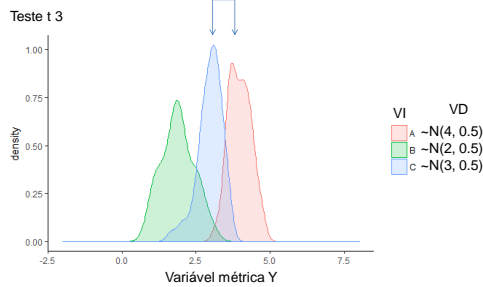


Introdução & Pressupostos Homocedasticidade & Médias ANOVA de um fator ANOVA com dois fatores ANOVA medidas repetidas

Teste t 2



Introdução & Pressupostos Homocedasticidade & Médias ANOVA de um fator ANOVA com dois fatores ANOVA medidas repetidas



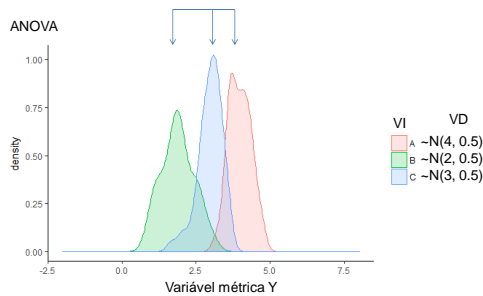
NA PRÁTICA

Cuidados:
 - Correção de Bonferroni
 Ao fazer 20 testes, 1/20 (0.05) pode resultar em dados significativos.
 Para tanto, fazer correção Bonferroni quando executar muitas análises (*Big Data*)

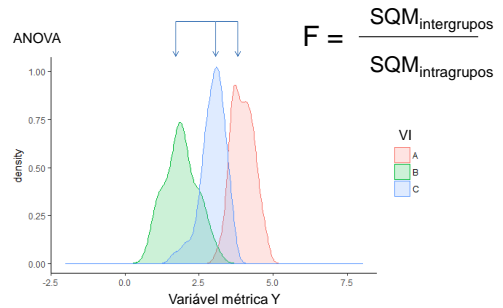


$$\alpha_{\text{corrigido}} = \frac{\alpha}{n \text{ de testes}}$$

Introdução & Pressupostos Homocedasticidade & Médias ANOVA de um fator ANOVA com dois fatores ANOVA medidas repetidas



Introdução & Pressupostos Homocedasticidade & Médias ANOVA de um fator ANOVA com dois fatores ANOVA medidas repetidas



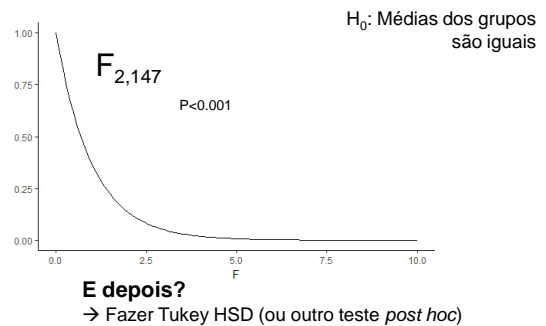
Introdução & Pressupostos Homocedasticidade & Médias ANOVA de um fator ANOVA com dois fatores ANOVA medidas repetidas

```
> anova(aov(y~A))
Analysis of Variance Table

Response: Variável métrica Y
VI      Df Sum Sq Mean Sq F value    Pr(>F)    
Residuals 147  42.288    0.288      204.17 < 2.2e-16 ***
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$$F = \frac{SQM_{\text{intergrupos}}}{SQM_{\text{intragrupos}}}$$

Introdução & Pressupostos Homocedasticidade & Médias ANOVA de um fator ANOVA com dois fatores ANOVA medidas repetidas



Introdução & Pressupostos	Homocedasticidade & Médias	ANOVA de um fator	ANOVA com dois fatores	ANOVA medidas repetidas
------------------------------	-------------------------------	----------------------	---------------------------	----------------------------

Teste Tukey HSD (Honest Significant Difference)

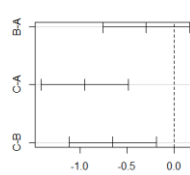
$$HSD = \frac{\bar{x}_a - \bar{x}_b}{\sqrt{\frac{MS_{within}}{n_{grupo}}}}$$

→ É realizado
APÓS ANOVA

Introdução & Pressupostos	Homocedasticidade & Médias	ANOVA de um fator	ANOVA com dois fatores	ANOVA medidas repetidas
------------------------------	-------------------------------	----------------------	---------------------------	----------------------------

As diferenças são significativas quando os
“intervalos” não estão encostados no eixo 0.

95% family-wise confidence level



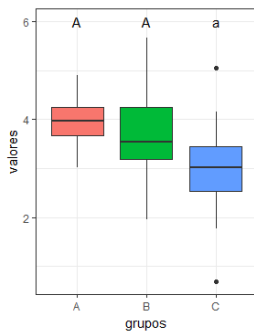
```
> TukeyHSD(anova_mod)
Tukey multiple comparisons of means
95% family-wise confidence level

Fit: aov(formula = valores ~ grupos, data = dados_1)

$grupos
      diff      lwr      upr    p.adj
B-A -0.2981098 -0.7668306  0.1666110  0.2791641
C-A -0.9520277 -1.4147486 -0.4893069  0.0001227
C-B -0.6539179 -1.1166388 -0.1911971  0.0031928

> plot(TukeyHSD(anova_mod))
>
```

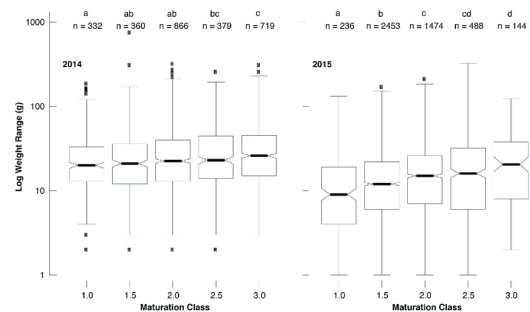
Introdução & Pressupostos	Homocedasticidade & Médias	ANOVA de um fator	ANOVA com dois fatores	ANOVA medidas repetidas
------------------------------	-------------------------------	----------------------	---------------------------	----------------------------



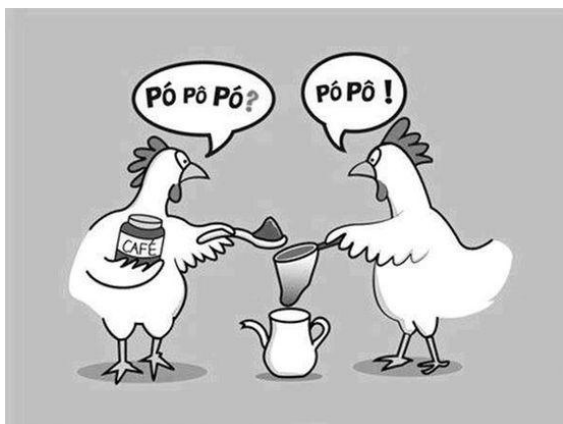
Escolhe-se um padrão de **letras diferentes** (ou tamanho de letras diferentes) para “agrupar” médias semelhantes.
→ Neste caso, grupos **A** e **B** são semelhantes, enquanto que o grupo **C** é diferente dos grupos **A** e **B**

Introdução & Pressupostos	Homocedasticidade & Médias	ANOVA de um fator	ANOVA com dois fatores	ANOVA medidas repetidas
------------------------------	-------------------------------	----------------------	---------------------------	----------------------------

EXEMPLO



Fonte: <http://doi.org/10.1371/journal.pone.0170375>



Introdução & Pressupostos	Homocedasticidade & Médias	ANOVA de um fator	ANOVA com dois fatores	ANOVA medidas repetidas
------------------------------	-------------------------------	----------------------	---------------------------	----------------------------

ETAPAS NECESSÁRIAS PARA EFETUAR UMA ANOVA

1. Verifique se os dados contínuos seguem uma distribuição normal
`shapiro.test(y)`
2. Verifique o pressuposto de homocedasticidade
`var.test(y ~ x)`
`bartlett.test(y ~ x)`
`levene.test(y ~ x)`
3. Variáveis são independentes?
+ **Número amostral semelhante**
+ **Amostragem suficiente**

Introdução & Pressupostos Homocedasticidade & Médias ANOVA de um fator ANOVA com dois fatores ANOVA medidas repetidas

```
> ANOVA_adit <- aov(len ~ factor(dose) + supp, data= dente)
> summary(ANOVA_adit)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
factor(dose)	2	2426.4	1213.2	82.81	< 2e-16 ***
supp	1	205.4	205.4	14.02	0.000429 ***
Residuals	56	820.4	14.7		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

$$Y \sim X_1 + X_2$$

Numérico Fator Fator

→ Esse é um modelo aditivo

Introdução & Pressupostos Homocedasticidade & Médias ANOVA de um fator ANOVA com dois fatores ANOVA medidas repetidas

```
Console ~/
> ANOVA_adit <- aov(len ~ factor(dose) + supp, data= dente)
> summary(ANOVA_adit)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
factor(dose)	2	2426.4	1213.2	82.81	< 2e-16 ***
supp	1	205.4	205.4	14.02	0.000429 ***
Residuals	56	820.4	14.7		

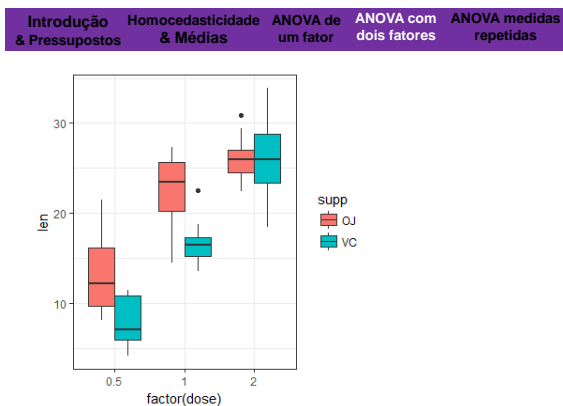
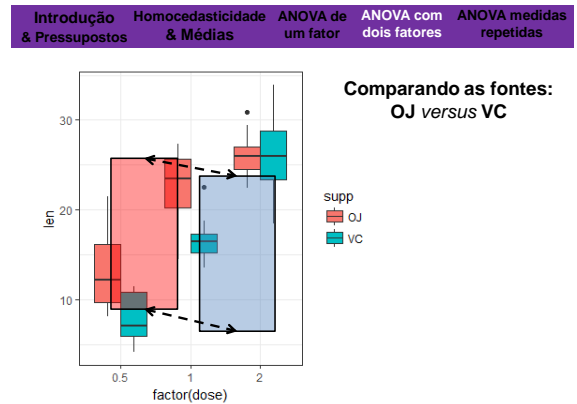
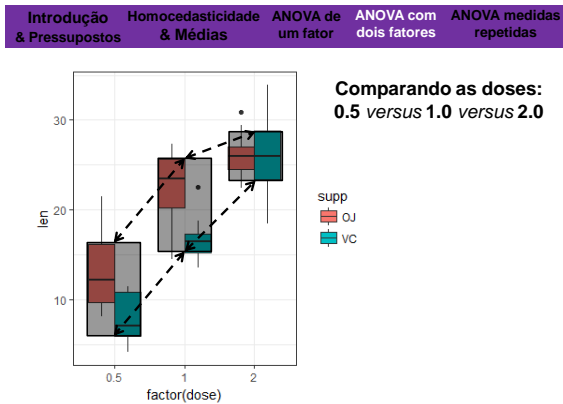
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
> ANOVA_LM_mod <- aov(len ~ dose + supp, data= dente)
> summary(ANOVA_LM_mod)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
dose	1	2224.3	2224.3	123.99	6.31e-16 ***
supp	1	205.3	205.3	11.45	0.0013 **
Residuals	57	1022.6	17.9		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

→ Se não estiver como fator, observe g.l.



Introdução & Pressupostos Homocedasticidade & Médias ANOVA de um fator ANOVA com dois fatores ANOVA medidas repetidas

supp	dose	supp:dose
OC	0.5	OC:0.5
	1.0	OC:1.0
	2.0	OC:2.0
VC	0.5	VC:0.5
	1.0	VC:1.0
	2.0	VC:2.0

→ Desenho balanceado?

```
table(dente$supp, dente$dose)
```

Introdução & Pressupostos	Homocedasticidade & Médias	ANOVA de um fator	ANOVA com dois fatores	ANOVA medidas repetidas
---------------------------	----------------------------	-------------------	------------------------	-------------------------

supp \ dose	0.5	1.0	2.0
VC	4.2	16.5	23.6
	11.5	16.5	18.5
	7.3	15.2	33.9
	5.8	17.3	25.5
	6.4	22.5	26.4
	10	17.3	32.5
	11.2	13.6	26.7
	11.2	14.5	21.5
OJ	5.2	18.8	23.3
	7	15.5	29.5
	15.2	19.7	25.5
	21.5	23.3	26.4
	17.6	23.6	22.4
	9.7	26.4	24.5
	14.5	20	24.8
	10	25.2	30.9
	8.2	25.8	26.4
	9.4	21.2	27.3
	16.5	14.5	29.4
	9.7	27.3	23

Introdução & Pressupostos	Homocedasticidade & Médias	ANOVA de um fator	ANOVA com dois fatores	ANOVA medidas repetidas
---------------------------	----------------------------	-------------------	------------------------	-------------------------

Soma dos Desvios Quadrados para o fator dose (colunas)

$$SSD_{colunas} = m \sum_j (\bar{x}_{\cdot j} - \bar{x}_{\cdot\cdot})^2$$

$\bar{x}_{\cdot j}$: Média da coluna j

$\bar{x}_{\cdot\cdot}$: Média geral

m : número de elementos para cada coluna j

Introdução & Pressupostos	Homocedasticidade & Médias	ANOVA de um fator	ANOVA com dois fatores	ANOVA medidas repetidas
---------------------------	----------------------------	-------------------	------------------------	-------------------------

supp \ dose	0.5	1.0	2.0
VC	4.2	16.5	23.6
	11.5	16.5	18.5
	7.3	15.2	33.9
	5.8	17.3	25.5
	6.4	22.5	26.4
	10	17.3	32.5
	11.2	13.6	26.7
	11.2	14.5	21.5
OJ	5.2	18.8	23.3
	7	15.5	29.5
	15.2	19.7	25.5
	21.5	23.3	26.4
	17.6	23.6	22.4
	9.7	26.4	24.5
	14.5	20	24.8
	10	25.2	30.9
	8.2	25.8	26.4
	9.4	21.2	27.3
	16.5	14.5	29.4
	9.7	27.3	23
$\bar{x}_{\cdot j}$	10.605	19.735	26.1

Introdução & Pressupostos	Homocedasticidade & Médias	ANOVA de um fator	ANOVA com dois fatores	ANOVA medidas repetidas
---------------------------	----------------------------	-------------------	------------------------	-------------------------

Soma dos Desvios Quadrados para o fator dose (colunas)

$$SSD_{colunas} = m \sum_j (\bar{x}_{\cdot j} - \bar{x}_{\cdot\cdot})^2$$

$$SSD_{colunas} = 20 \times [(10.605 - 18.81)^2 + (19.735 - 18.81)^2 + (26.1 - 18.81)^2]$$

$$SSD_{colunas} = 2426.43$$

gl = (classes nas colunas-1)

Introdução & Pressupostos	Homocedasticidade & Médias	ANOVA de um fator	ANOVA com dois fatores	ANOVA medidas repetidas
---------------------------	----------------------------	-------------------	------------------------	-------------------------

Soma dos Desvios Quadrados para o fator supp (linhas)

$$SSD_{linhas} = n \sum_i (\bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot})^2$$

$\bar{x}_{i\cdot}$: Média da linha i

$\bar{x}_{\cdot\cdot}$: Média geral

n : número de elementos para cada linha i

Introdução & Pressupostos	Homocedasticidade & Médias	ANOVA de um fator	ANOVA com dois fatores	ANOVA medidas repetidas
---------------------------	----------------------------	-------------------	------------------------	-------------------------

supp \ dose	0.5	1.0	2.0	$\bar{x}_{i\cdot}$
VC	4.2	16.5	23.6	16.963
	11.5	16.5	18.5	
	7.3	15.2	33.9	
	5.8	17.3	25.5	
	6.4	22.5	26.4	
	10	17.3	32.5	
	11.2	13.6	26.7	
	11.2	14.5	21.5	
OJ	5.2	18.8	23.3	20.663
	7	15.5	29.5	
	15.2	19.7	25.5	
	21.5	23.3	26.4	
	17.6	23.6	22.4	
	9.7	26.4	24.5	
	14.5	20	24.8	
	10	25.2	30.9	
	8.2	25.8	26.4	
	9.4	21.2	27.3	
	16.5	14.5	29.4	
	9.7	27.3	23	

Introdução & Pressupostos	Homocedasticidade & Médias	ANOVA de um fator	ANOVA com dois fatores	ANOVA medidas repetidas
---------------------------	----------------------------	-------------------	------------------------	-------------------------

Soma dos Desvios Quadrados para o fator *supp* (linhas)

$$SSD_{linhas} = n \sum_i (\bar{x}_{i\bullet} - \bar{x}_{\bullet\bullet})^2$$

$$SSD_{linhas} = 30 \times [(16.963 - 18.81)^2 + (20.663 - 18.81)^2]$$

$$SSD_{linhas} = 205.35$$

$$gl = (\text{classes nas linhas} - 1)$$

Introdução & Pressupostos	Homocedasticidade & Médias	ANOVA de um fator	ANOVA com dois fatores	ANOVA medidas repetidas
---------------------------	----------------------------	-------------------	------------------------	-------------------------

Soma dos Desvios Quadrados para a interação entre fator *supp* (linhas) e fator *dose* (colunas)

$$SSD_{int} = \sum_i \sum_j (\bar{x}_{ij} - \bar{x}_{i\bullet} - \bar{x}_{\bullet j} + \bar{x}_{\bullet\bullet})^2$$

$$gl = (\text{linhas} - 1) * (\text{colunas} - 1)$$

Introdução & Pressupostos	Homocedasticidade & Médias	ANOVA de um fator	ANOVA com dois fatores	ANOVA medidas repetidas
---------------------------	----------------------------	-------------------	------------------------	-------------------------

supp \ dose	0.5	1.0	2.0
VC	4.2 11.5 7.3 5.8 6.4 10 11.2 5.2 7	16.5 16.5 15.2 17.3 22.5 17.3 13.6 14.5 18.8 15.5	23.6 18.5 33.9 25.5 26.4 32.5 26.7 21.5 23.3 29.5
\bar{x}_{ij}	7.98	16.77	26.14
OJ	15.2 21.5 17.6 9.7 14.5 10 8.2 9.4 16.5 9.7	19.7 23.3 23.6 26.4 20 25.2 25.8 21.2 14.5 27.3	25.5 26.4 22.4 24.5 24.8 30.9 26.4 27.3 29.4 23
\bar{x}_{ij}	13.23	22.7	26.06

Introdução & Pressupostos	Homocedasticidade & Médias	ANOVA de um fator	ANOVA com dois fatores	ANOVA medidas repetidas
---------------------------	----------------------------	-------------------	------------------------	-------------------------

Soma dos Desvios Quadrados para a interação entre fator *supp* (linhas) e fator *dose* (colunas)

$$SSD_{int} = \sum_i \sum_j (\bar{x}_{ij} - \bar{x}_{i\bullet} - \bar{x}_{\bullet j} + \bar{x}_{\bullet\bullet})^2$$

supp \ dose	0.5	1.0	2.0	
VC				
OJ				

Introdução & Pressupostos	Homocedasticidade & Médias	ANOVA de um fator	ANOVA com dois fatores	ANOVA medidas repetidas
---------------------------	----------------------------	-------------------	------------------------	-------------------------

Soma dos Desvios Quadrados para a interação entre fator *supp* (linhas) e fator *dose* (colunas)

$$SSD_{int} = \sum_i \sum_j (\bar{x}_{ij} - \bar{x}_{i\bullet} - \bar{x}_{\bullet j} + \bar{x}_{\bullet\bullet})^2$$

supp \ dose	0.5	1.0	2.0	
VC				
OJ				
				18.81

Introdução & Pressupostos	Homocedasticidade & Médias	ANOVA de um fator	ANOVA com dois fatores	ANOVA medidas repetidas
---------------------------	----------------------------	-------------------	------------------------	-------------------------

Soma dos Desvios Quadrados para a interação entre fator *supp* (linhas) e fator *dose* (colunas)

$$SSD_{int} = \sum_i \sum_j (\bar{x}_{ij} - \bar{x}_{i\bullet} - \bar{x}_{\bullet j} + \bar{x}_{\bullet\bullet})^2$$

supp \ dose	0.5	1.0	2.0	$\bar{x}_{i\bullet}$
VC				16.96
OJ				20.66
$\bar{x}_{\bullet j}$	10.61	19.73	26.1	18.81

Introdução & Pressupostos	Homocedasticidade & Médias	ANOVA de um fator	ANOVA com dois fatores	ANOVA medidas repetidas
------------------------------	-------------------------------	----------------------	---------------------------	----------------------------

Soma dos Desvios Quadrados para a interação entre fator *supp* (linhas) e fator *dose* (colunas)

$$SSD_{\text{int}} = \sum_i \sum_j (x_{ij} - \bar{x}_{i\bullet} - \bar{x}_{\bullet j} + \bar{x}_{\bullet\bullet})^2$$

supp	dose	0.5	1.0	2.0	$\bar{x}_{i\bullet}$
VC					16.96
x_{ij}		7.98	16.77	26.14	
OJ					20.66
x_{ij}		13.23	22.7	26.06	
$\bar{x}_{\bullet j}$		10.61	19.73	26.1	18.81

Introdução & Pressupostos	Homocedasticidade & Médias	ANOVA de um fator	ANOVA com dois fatores	ANOVA medidas repetidas
------------------------------	-------------------------------	----------------------	---------------------------	----------------------------

Soma dos Desvios Quadrados para a interação entre fator *supp* (linhas) e fator *dose* (colunas)

$$SSD_{\text{int}} = \sum_i \sum_j (x_{ij} - \bar{x}_{i\bullet} - \bar{x}_{\bullet j} + \bar{x}_{\bullet\bullet})^2$$

supp	dose	0.5	1.0	2.0	$\bar{x}_{i\bullet}$
VC		SSD_{11}	SSD_{12}	SSD_{13}	16.96
x_{ij}		7.98	16.77	26.14	
OJ		SSD_{21}	SSD_{22}	SSD_{23}	20.66
x_{ij}		13.23	22.7	26.06	
$\bar{x}_{\bullet j}$		10.61	19.73	26.1	18.81

Introdução & Pressupostos	Homocedasticidade & Médias	ANOVA de um fator	ANOVA com dois fatores	ANOVA medidas repetidas
------------------------------	-------------------------------	----------------------	---------------------------	----------------------------

Soma dos Desvios Quadrados para a interação entre fator *supp* (linhas) e fator *dose* (colunas)

$$SSD_{\text{int}} = \sum_i \sum_j (x_{ij} - \bar{x}_{i\bullet} - \bar{x}_{\bullet j} + \bar{x}_{\bullet\bullet})^2$$

supp	dose	0.5	1.0	2.0	$\bar{x}_{i\bullet}$
VC		(7.98 - 16.96 - 10.61 + 18.81) ²	(16.77 - 16.96 - 19.73 + 18.81) ²	(26.14 - 16.96 - 26.01 + 18.81) ²	16.96
x_{ij}		7.98	16.77	26.14	
OJ		(13.23 - 20.66 - 10.61 + 18.81) ²	(22.7 - 20.66 - 19.73 + 18.81) ²	(26.06 - 20.66 - 26.1 + 18.81) ²	20.66
x_{ij}		13.23	22.7	26.06	
$\bar{x}_{\bullet j}$		10.61	19.73	26.1	18.81

Introdução & Pressupostos	Homocedasticidade & Médias	ANOVA de um fator	ANOVA com dois fatores	ANOVA medidas repetidas
------------------------------	-------------------------------	----------------------	---------------------------	----------------------------

Soma dos Desvios Quadrados para a interação entre fator *supp* (linhas) e fator *dose* (colunas)

$$SSD_{\text{int}} = \sum_i \sum_j (x_{ij} - \bar{x}_{i\bullet} - \bar{x}_{\bullet j} + \bar{x}_{\bullet\bullet})^2$$

supp	dose	0.5	1.0	2.0	
VC		10*0.601	10*1.243	10*3.572	
OJ		10*0.601	10*1.243	10*3.5721	

Introdução & Pressupostos	Homocedasticidade & Médias	ANOVA de um fator	ANOVA com dois fatores	ANOVA medidas repetidas
------------------------------	-------------------------------	----------------------	---------------------------	----------------------------

Soma dos Desvios Quadrados para a interação entre fator *supp* (linhas) e fator *dose* (colunas)

$$SSD_{\text{int}} = \sum_i \sum_j (x_{ij} - \bar{x}_{i\bullet} - \bar{x}_{\bullet j} + \bar{x}_{\bullet\bullet})^2$$

$$SSD_{\text{int}} = 108.3$$

$$gl = (\text{linhas}-1) * (\text{colunas}-1)$$

Introdução & Pressupostos	Homocedasticidade & Médias	ANOVA de um fator	ANOVA com dois fatores	ANOVA medidas repetidas
------------------------------	-------------------------------	----------------------	---------------------------	----------------------------

Soma dos Desvios Quadrados TOTAL

$$SSD_{\text{TOTAL}} = \sum (x_{ij} - \bar{x}_{\bullet\bullet})^2$$

$$SSD_{\text{int}} = 3452.21$$

$$gl = n - 1$$

Introdução & Pressupostos	Homocedasticidade & Médias	ANOVA de um fator	ANOVA com dois fatores	ANOVA medidas repetidas
---------------------------	----------------------------	-------------------	------------------------	-------------------------

Soma dos Desvios Quadrados Resíduos

$$SSD_{TOTAL} = SSD_{resíduos} + SSD_{linhas} + SSD_{colunas} + SSD_{int}$$

$$345221 = SSD_{resíduos} + 2426.4 + 205.35 + 108.3$$

$$SSD_{resíduos} = 712.1$$

$$gl = l.c.(n'-1)$$

Introdução & Pressupostos	Homocedasticidade & Médias	ANOVA de um fator	ANOVA com dois fatores	ANOVA medidas repetidas
---------------------------	----------------------------	-------------------	------------------------	-------------------------

QUADRO ANOVA 2-FATORES

Fonte da variação	Soma dos Quadrados	Graus de liberdade	Quadrados Médios QM (variâncias)	F
Linhas	SSD _{linhas}	Linhas-1	SSD _{linhas} /Linhas-1	QM _{lin} /QM _{res}
Colunas	SSD _{colunas}	colunas-1	SSD _{colunas} /col-1	QM _{col} /QM _{res}
Linhas:Colunas	SSD _{inter}	(l-1)(c-1)	SSD _{inter} /(l-1)(c-1)	QM _{int} /QM _{res}
Resíduos	SSD _{resíduos}	l.c.(n'-1)	SSD _{inter} /l.c.(n'-1)	
TOTAL	SSD _{total}	n-1		

Introdução & Pressupostos	Homocedasticidade & Médias	ANOVA de um fator	ANOVA com dois fatores	ANOVA medidas repetidas
---------------------------	----------------------------	-------------------	------------------------	-------------------------

QUADRO ANOVA 2-FATORES

Fonte da variação	Soma dos Quadrados	Graus de liberdade	Quadrados Médios QM (variâncias)	F
supp	205.4	2-1	205.4	92.0
dose	2426.4	3-1	1213.2	15.6
supp:dose	108.3	(2-1)(3-1)	54.2	4.1
Resíduos	712.1	2.3(10-1)	13.2	
TOTAL	3452.2	60-1		

Introdução & Pressupostos	Homocedasticidade & Médias	ANOVA de um fator	ANOVA com dois fatores	ANOVA medidas repetidas
---------------------------	----------------------------	-------------------	------------------------	-------------------------

```
> ANOVA_int <- aov(len ~ factor(dose) + supp + factor(dose):supp, data= dente)
> summary(ANOVA_int)
              Df Sum Sq Mean Sq F value    Pr(>F)
factor(dose)    2  2426.4   1213.2    92.000 < 2e-16 ***
supp             1   205.4    205.4   15.572 0.000231 ***
factor(dose):supp 2   108.3     54.2    4.107 0.021860 *
Residuals       54   712.1     13.2
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> ANOVA_int <- aov(len ~ factor(dose)*supp, data= dente)
> summary(ANOVA_int)
              Df Sum Sq Mean Sq F value    Pr(>F)
factor(dose)    2  2426.4   1213.2    92.000 < 2e-16 ***
supp             1   205.4    205.4   15.572 0.000231 ***
factor(dose):supp 2   108.3     54.2    4.107 0.021860 *
Residuals       54   712.1     13.2
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
```

→ Com interação entre fatores

→ Quando a interação **não é significativa**, levar em consideração somente **modelo aditivo**

Introdução & Pressupostos	Homocedasticidade & Médias	ANOVA de um fator	ANOVA com dois fatores	ANOVA medidas repetidas
---------------------------	----------------------------	-------------------	------------------------	-------------------------

INTERPRETAÇÃO

- A dose recebida (dose) influencia nas médias do comprimento dos dentes

$$F_{2,54} = 92, \text{ p-valor} < 0.0001$$

.

- A fonte da vitamina C (supp) recebida influencia nas médias do comprimento dos dentes

$$F_{1,54} = 15.57, \text{ p-valor} < 0.0001$$

- A relação entre dose e comprimento dos dentes é influenciada pela fonte de vitamina C

$$F_{2,54} = 4.11, \text{ p-valor} < 0.05$$

Introdução & Pressupostos	Homocedasticidade & Médias	ANOVA de um fator	ANOVA com dois fatores	ANOVA medidas repetidas
---------------------------	----------------------------	-------------------	------------------------	-------------------------

```
> TukeyHSD(ANOVA_int)
Tukey multiple comparisons of means
95% family-wise confidence level

Fit: aov(formula = len ~ factor(dose) * supp, data = dente)

$`factor(dose):supp`
              diff            twr            upr      p adj
1-0.5    9.130    6.362488    11.897512 0.0e+00
2-0.5   15.495   12.727488   18.262512 0.0e+00
2-1      6.365    3.597488    9.132512 2.7e-06

$supp
              diff            twr            upr      p adj
VC-0J   -3.7    -5.579828   -1.820172 0.0002312

$`factor(dose):supp`
              diff            twr            upr      p adj
1:0J-0.5:0J    9.47    4.671876   14.2681238 0.0000046
2:0J-0.5:0J   12.83    8.031876   17.6281238 0.0000000
0.5:VC-0.5:0J  -5.25   -10.048124   -0.4518762 0.0242521
1:VC-0.5:0J    3.54   -1.258124    8.3381238 0.2640208
2:VC-0.5:0J   12.91    8.111876   17.7081238 0.0000000
2:0J-1:0J      3.36   -1.438124    8.1581238 0.3187361
0.5:VC-1:0J   -14.72   -19.518124   -9.9218762 0.0000000
1:VC-1:0J     -5.93   -10.728124   -1.1318762 0.0073910
2:VC-1:0J      3.44   -1.358124    8.2381238 0.2936430
0.5:VC-2:0J   -16.08   -22.878124   -13.2818762 0.0000000
1:VC-2:0J     -9.29   -14.088124   -4.4918762 0.0000069
2:VC-2:0J      0.08   -4.718124    4.8781238 1.0000000
1:VC-0.5:VC     8.79    3.991876   13.5881238 0.0000210
2:VC-0.5:VC    18.16   13.361876   22.9581238 0.0000000
2:VC-1:VC      9.37    4.571876   14.1681238 0.0000058
```


Introdução & Pressupostos **Homocedasticidade & Médias** **ANOVA de um fator** **ANOVA com dois fatores** **ANOVA medidas repetidas**

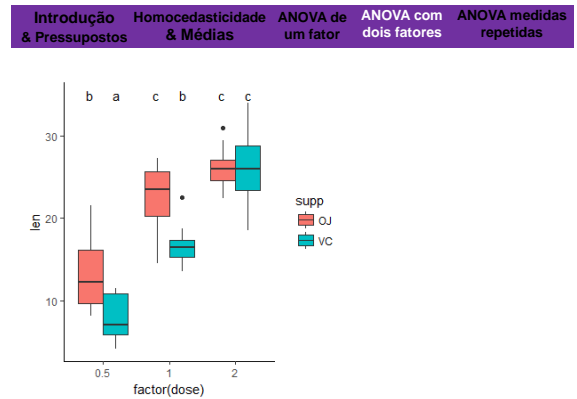
```
if(!require(lme4)){install.packages("lme4")}
model <- lm(len ~ factor(dose)*supp, data= dente)
marginal <- lme4::lsmeans(model,
  pairwise ~ factor(dose):supp,
  adjust="tukey")
```

Console

```
> cld(marginal,
+ alpha=0.05,      ## Nível de significância da diferença dos grupos
+ Letters=letters,  ## Usa letras caixa baixa para cada grupo
+ adjust="tukey")    ## Comparação ajustada de Tukey
```

dose	supp	lsmean	SE	df	lower.CL	upper.CL	group
0.5	VC	7.98	1.148353	54	4.843907	11.11609	a
0.5	OJ	13.23	1.148353	54	10.093907	16.36609	b
1.0	VC	16.77	1.148353	54	13.633907	19.90609	b
1.0	OJ	22.70	1.148353	54	19.563907	25.83609	c
2.0	OJ	26.06	1.148353	54	22.923907	29.19609	c
2.0	VC	26.14	1.148353	54	23.003907	29.27609	c

Confidence level used: 0.95
 Conf-level adjustment: sidak method for 6 estimates
 P value adjustment: tukey method for comparing a family of 6 estimates
 Significance level used: alpha = 0.05



Introdução & Pressupostos **Homocedasticidade & Médias** **ANOVA de um fator** **ANOVA com dois fatores** **ANOVA medidas repetidas**

ANCOVA
ANÁLISE de COVARIÂNCIA

```
> ANCOVA_ <- aov(len ~ dose*supp, data= dente)
> summary(ANCOVA_)
```

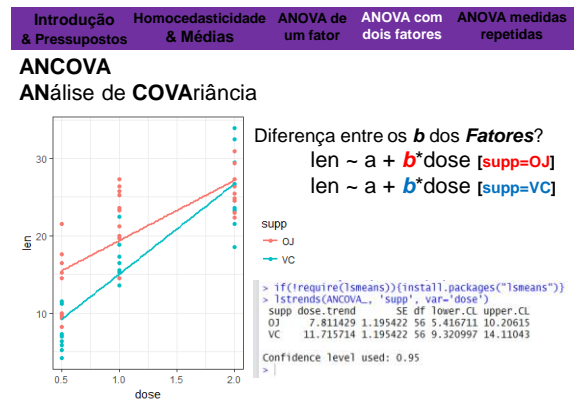
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
dose	1	2224.3	2224.3	133.415	< 2e-16 ***
supp	1	205.3	205.3	12.317	0.000894 ***
dose:supp	1	88.9	88.9	5.333	0.024631 *
Residuals	56	933.6	16.7		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Y ~ X1*cov(X2)

Numérico Fator Numérico

+ DISCIPLINA 11



Introdução & Pressupostos **Homocedasticidade & Médias** **ANOVA de um fator** **ANOVA com dois fatores** **ANOVA medidas repetidas**

E SE OS FATORES ESTÃO DESBALANCEADOS?
 → Utiliza-se Soma dos Quadrados de diferentes Tipos (*sensu* SAS)

Tipo I
 O cálculo da Soma dos Quadrados é sequencial. Primeiro testa o efeito do fator A, depois do fator B e por último o fator A:B. Portanto, sofre com duas influências: (i) ordem dos fatores postos no modelo e (ii) se os fatores estão desbalanceados.

Tipo II
 Testa primeiro a interação entre os fatores A:B, depois é que analisa separadamente os fatores A e B. Esse é mais indicado (mais poderoso) quando não há interação entre os fatores (interação não é significativa)

Tipo III
 Válida para quando há interação significativa e os fatores estão desbalanceados. Normalmente, quando a interação é significativa, não é interessante ver o efeito dos fatores separadamente.

Ver também:
<https://mcfromnz.wordpress.com/2011/03/02/anova-type-iii-ss-explained/>
<https://stats.stackexchange.com/questions/60362/choice-between-type-i-type-ii-or-type-iii-anova>

Introdução & Pressupostos **Homocedasticidade & Médias** **ANOVA de um fator** **ANOVA com dois fatores** **ANOVA medidas repetidas**

E SE OS FATORES ESTÃO DESBALANCEADOS?

```
> if(!require(car)){install.packages("car")}
> set.seed(32)
> meus_dentes <- sample_n(dente, 32)
> table(meus_dentes$supp, meus_dentes$dose)
```

	0.5	1	2
OJ	8	7	4
VC	2	6	5

```
>
> minha_anova <- aov(len ~ supp * factor(dose), data = meus_dentes)
> Anova(minha_anova, type = "III")
Anova Table (Type III tests)
```

	Sum Sq	Df	F value	Pr(>F)
(Intercept)	1277.65	1	111.8134	6.546e-11 ***
supp	40.60	1	3.5533	0.07065 .
factor(dose)	601.38	2	26.3150	5.649e-07 ***
supp:factor(dose)	51.33	2	2.2461	0.12595
Residuals	297.09	26		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Introdução & Pressupostos	Homocedasticidade & Médias	ANOVA de um fator	ANOVA com dois fatores	ANOVA medidas repetidas
------------------------------	-------------------------------	----------------------	---------------------------	----------------------------

E SE OS FATORES ESTÃO DESBALANCEADOS?

```
> Anova(minha_anova, type = "II")
Anova Table (Type II tests)

Response: len
          Sum Sq Df F value    Pr(>F)    
supp      124.79  1 10.9213  0.002775 **
factor(dose) 1051.97 2 46.0316 2.865e-09 ***
supp:factor(dose) 51.33 2  2.2461  0.125948
Residuals    297.09 26
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
```

Introdução & Pressupostos	Homocedasticidade & Médias	ANOVA de um fator	ANOVA com dois fatores	ANOVA medidas repetidas
------------------------------	-------------------------------	----------------------	---------------------------	----------------------------

EXERCÍCIO

```
data("CO2")
?CO2
```

1. Testar se Treatment e Type influenciam em uptake.
2. Verificar se os dados seguem a distribuição normal (*)
3. Verificar se os dados são homocedásticos
4. Verificar se o experimento é balanceado
5. Executar ANOVA
6. Executar teste *post hoc*?
7. Concluir



Introdução & Pressupostos	Homocedasticidade & Médias	ANOVA de um fator	ANOVA com dois fatores	ANOVA medidas repetidas
------------------------------	-------------------------------	----------------------	---------------------------	----------------------------

EXERCÍCIO

```
data("CO2")
?CO2
```

1. Testar se Type influencia na relação entre conc e uptake.
2. Executar teste *post hoc*?



Introdução & Pressupostos	Homocedasticidade & Médias	ANOVA de um fator	ANOVA com dois fatores	ANOVA medidas repetidas
------------------------------	-------------------------------	----------------------	---------------------------	----------------------------

Prós

- Custa menos (precisa de menos sujeitos)
- Maior poder estatístico

Contra:

- Não contabiliza "Error term" (em nome do poder estatístico)
- Precisa estar em ordem ("Tempo1", "Tempo2" ...)
- Valores faltantes

+ DISCIPLINA 11

Introdução & Pressupostos	Homocedasticidade & Médias	ANOVA de um fator	ANOVA com dois fatores	ANOVA medidas repetidas
------------------------------	-------------------------------	----------------------	---------------------------	----------------------------

Viola o pressuposto de independência

→ Um sujeito é medido mais do que uma vez

Pressuposto de esfericidade:

→ Relação entre os pares das condições experimentais é similar: nível de dependência é praticamente igual

$$\text{Variância A-C} = \text{Variância A-B} = \text{Variância B-C}$$

Esfericidade: A estrutura da variância-covariância das medidas repetidas segue um determinado padrão.

Introdução & Pressupostos	Homocedasticidade & Médias	ANOVA de um fator	ANOVA com dois fatores	ANOVA medidas repetidas
---------------------------	----------------------------	-------------------	------------------------	-------------------------

Organização dos dados

Dados "WIDE" ⇔ Dados "LONG"

ID	Fator	Temp1	Temp2
1	Trata	0.1	0.2
2	Control	1.1	1.2
...			
n	FatorX	5.1	5.2

ID	Fator	Tempo	Valor
1	Trata	1	0.1
1	Trata	2	0.2
2	Control	1	1.1
2	Control	2	1.2
...			
n	FatorX	2	5.2

Introdução & Pressupostos	Homocedasticidade & Médias	ANOVA de um fator	ANOVA com dois fatores	ANOVA medidas repetidas
---------------------------	----------------------------	-------------------	------------------------	-------------------------

EXEMPLO

```
data("OBrienKaiser")
?OBrienKaiser
```

- 16 Indivíduos: Homens (M) e Mulheres (F)
- 3 Tempos: Pré, Pós e "Follow-up"
- 3 Tratamentos: Controle, A e B
- 5 Medidas em cada Tempo



Introdução & Pressupostos	Homocedasticidade & Médias	ANOVA de um fator	ANOVA com dois fatores	ANOVA medidas repetidas
---------------------------	----------------------------	-------------------	------------------------	-------------------------

EXEMPLO

```
if(!require(tidyr)){install.packages("tidyr")}
dados_wide <- subset(OBrienKaiser,
                     select = c(treatment,
                                gender,
                                pre.1,
                                post.1,
                                fup.1))

dados_wide$ID <- 1:16

dados_long <- gather(dados_wide,
                     key = tempo,
                     value = valor_1,
                     -c(ID, treatment, gender))

dados_long$tempo <- ordered(dados_long$tempo,
                             levels = c("pre", "post", "fup"))
```



Introdução & Pressupostos	Homocedasticidade & Médias	ANOVA de um fator	ANOVA com dois fatores	ANOVA medidas repetidas
---------------------------	----------------------------	-------------------	------------------------	-------------------------

EXEMPLO

Dados "LONG"

```
summary(aov(VD ~VI+ Error(ID/VI),
             data=dados_long))
```

Como são medidas repetidas (de um mesmo ID), você usa esse termo de "Error", dizendo que os sujeitos (ID) estão sendo medidos repetidamente ao longo dos fatores (VI) para a variável contínua (VD)



Introdução & Pressupostos	Homocedasticidade & Médias	ANOVA de um fator	ANOVA com dois fatores	ANOVA medidas repetidas
---------------------------	----------------------------	-------------------	------------------------	-------------------------

EXEMPLO

Dados "WIDE"

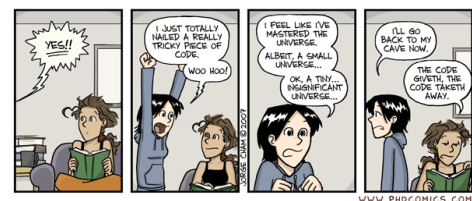
```
modelo <- lm(cbind(VD.1, VD.2, VD.3) ~ 1,
             data = dados_wide)

dados_wide$VI <- factor(dados_wide$VI)

iDado <- data.frame(VarInd = levels(dados_wide$VI))

AnovaMR <- car::Anova(modelo,
                       idata = iDado,
                       idesign = ~VarInd)

summary(AnovaMR, multivariate = FALSE,
        univariate = TRUE)
```



Análises de Variâncias

TESTES NÃO-PARAMÉTRICOS

Kruskal-Wallis

→ É alternativa não-paramétrica para o teste ANOVA para um fator (one-way ANOVA).
→ Semelhante ao Teste U de Wilcoxon pois também utiliza **ranking**.

Características

- Análise de variância não paramétrica
- 3 ou + grupos independentes
- Hipótese: As distribuições de todos os grupos são iguais,
- Hipótese: As medianas de todos os grupos são iguais
- Insensível a outliers
- Os grupos não precisam ter o mesmo tamanho

Kruskal-Wallis

$$H = \frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(N+1)$$

n_i : número de observações no grupo i

N : número total de observações em todos os k grupos

R_i : soma dos rankings das n_i observações no grupo i

Kruskal-Wallis

Cuidados

- Se a distribuição for normal é melhor usar o teste ANOVA de um critério (one-way)
- Precisa ter 4 ou mais elementos na amostra de cada grupo
- Se tiver só 2 grupos use o Mann-Whitney

Kruskal-Wallis

Numbers of Flies/m ³ of Foliage		
Herbs	Shrubs	Trees
14.0 (15)	8.4 (11)	6.9 (8)
12.1 (14)	5.1 (2)	7.3 (9)
9.6 (12)	5.5 (4)	5.8 (5)
8.2 (10)	6.6 (7)	4.1 (1)
10.2 (13)	6.3 (6)	5.4 (3)
$n_1 = 5$	$n_2 = 5$	$n_3 = 5$
$R_1 = 64$	$R_2 = 30$	$R_3 = 26$

$$\begin{aligned}
 N &= 5 + 5 + 5 = 15 \\
 H &= \frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(N+1) \\
 &= \frac{12}{15(16)} \left[\frac{64^2}{5} + \frac{30^2}{5} + \frac{26^2}{5} \right] - 3(16) \\
 &= \frac{12}{240} [1134.400] - 48 \\
 &= 56.720 - 48 \\
 &= 8.720
 \end{aligned}$$

Calculado ←

Critico (tabelado) ← $H_{0.05,5,5} = 5.780$
Reject H_0

Fonte: Zar, J.H. 2010. *Biostatistical Analysis*. 5th ed. Ed. Pearson.

Kruskal-Wallis

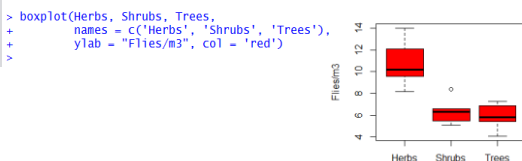
```

> Herbs <- c(14, 12.1, 9.6, 8.2, 10.2)
> Shrubs <- c(8.4, 5.1, 5.5, 6.6, 6.3)
> Trees <- c(6.9, 7.3, 5.8, 4.1, 5.4)
> flies <- data.frame(local=c(rep('Herbs', 5),
+                             rep('Shrubs', 5),
+                             rep('Trees', 5)),
+                     obs=c(Herbs, Shrubs, Trees))
> kruskal.test(obs~local, flies)

Kruskal-Wallis rank sum test

data:  obs by local
Kruskal-Wallis chi-squared = 8.72, df = 2, p-value = 0.01278

```



Friedman

→ Alternativa para ANOVA com dois fatores (two-way ANOVA)
→ Equivalente ao **Teste de Sinais**, onde testa pares de + ou - dentro de cada par.
→ É menos sensível que o teste de sinais de Wilcoxon

```

> sample_n(CO2, 5)
Grouped Data: uptake ~ conc | Plant
  Plant  Type  Treatment conc uptake
77  MC2 Mississippi chilled 1000  14.4
69  MC1 Mississippi chilled  675  22.2
37   QC3  Quebec    chilled  175  21.0
63  Mn3 Mississippi nonchilled 1000  27.8
42   QC3  Quebec    chilled  1000  41.4
> friedman.test(uptake~conc|Plant, data = CO2)

Friedman rank sum test

data:  uptake and conc and Plant
Friedman chi-squared = 59.677, df = 6, p-value = 5.236e-11

```